CENTRE D'ETUDES
ET DE RECHERCHES
SUR LE DEVELOPPEMENT
INTERNATIONAL

Document de travail de la série

Etudes et Documents

E 2011.17

Transboundary Pollution in China:
A Study of the Location Choice of Polluting Firms in Hebei Province

Chloé Duvivier

Hang Xiong

May, 2011

CERDI
65 BD. F. MITTERRAND
63000 CLERMONT FERRAND - FRANCE
TEL. 04 73 17 74 00
FAX 04 73 17 74 28
www.cerdi.org

Les auteurs

Chloé Duvivier
PhD student, Clermont Université, Université d'Auvergne, CNRS, UMR 6587, Centre d'Etudes et de Recherches sur le Développement International (CERDI), F-63009 Clermont-Ferrand, France
Email : duvivierchloe@gmail.com

Hang Xiong
PhD student, Clermont Université, Université d'Auvergne, CNRS, UMR 6587, Centre d'Etudes et de Recherches sur le Développement International (CERDI), F-63009 Clermont-Ferrand, France
Email : m.audibert@u-clermont1.fr

Corresponding author: C. Duvivier, CERDI-CNRS. E-mail: duvivierchloe@gmail.com

La série des *Etudes et Documents* du CERDI est consultable sur le site :

http://www.cerdi.org/ed

Directeur de la publication : Patrick Plane

Directeur de la rédaction : Catherine Araujo Bonjean

Responsable d'édition : Annie Cohade

ISSN : 2114-7957

Abstract

In this paper we study whether or not transboundary pollution problems exist in China. To do so, we estimate whether, within Hebei province, polluting firms are more likely to set up in border counties than in interior ones. For this purpose, we use the lists of polluting firms published annually by the Ministry of Environmental Protection of China and by the Environmental Protection Bureau of Hebei province. To ensure the robustness of our results, several measures of the variable of interest are constructed from GIS data. The estimations of a count-data model allow us to conclude that border counties are more attractive destinations for polluting firms than counties located within the province. Moreover, it appears that this effect has strengthened over time.

Keywords: Transboundary pollution, firm location choice, environmental regulations, China

JEL codes: Q01, Q56

Acknowledgements

We would like to thank Eric Langlois and Olivier Santoni for their help with GIS data. We also would like to thank Jean-Louis Combes, Johanna Choumert, Christian Ebeke, Sandra Poncet and the participants in various seminars for helpful comments.

1 Introduction

Since the beginning of the reforms, China has achieved spectacular economic performances. The high and sustained economic growth allowed to reduce the proportion of poor people in the population from 65% to 10%, that is a fall of nearly half a billion in the number of poor people (World Bank, 2009). However, dramatic environmental degradation have accompanied these changes. The level of pollution is now threatening economic development and public health in the country. The total cost of air and water pollution in China amounts to 5.8% of GDP. Among these costs, health costs are the most important, representing 4.3% of the Chinese GDP (World Bank, 2007). Nowadays, cancer has become the first cause of death in China, explaining respectively 25% and 21% of deaths in urban and rural areas. (Ministry of Health, 2008[1]).

If the country's environmental problems are extremely severe, they are also very complex. In China, the environmental policy is decentralized: the central government sets the standards to reach and the local governments monitor and impose sanctions on polluters. Several studies have shown that if a decentralized environmental policy allows for greater flexibility as well as better information and adaptation to the local context, it can also lead to strategic behavior problems and to an excess of pollution at the borders (Oates, 2001 ; Sigman, 2005). Transboundary pollution problems can arise from two phenomena: free-riding and pollution-havens. On the one hand, local regulators prefer polluting firms to set up near the borders. This allows them to enjoy the economic benefits (taxes and employment) associated with the presence of industries while exporting some of the pollution generated to the neighboring provinces (free-riding hypothesis). On the other hand, at borders, there are significant discontinuities in environmental regulation. A firm can set up near a border in order to evade strict environmental regulations while continuing to benefit from the market of neighboring regions with stricter regulations (pollution-haven hypothesis). As a result, transboundary pollution problems are often cited as one very important drawback of decentralization by its opponents.

In the context of China where the environmental policy is decentralized and where it exerts an important effect on polluting firms (Dasgupta *et al.*, 2001, Wang and Wheeler, 2003, 2005, Wang, 2002), we expect that transboundary pollution problems exist. Moreover, several environmental conflicts between residents of a province and polluters from the neighboring province have already attracted great attention. The Huai River that runs through four provinces (Henan, Shandong, Anhui and Jiangsu) illustrates this. On several occasions, the downstream provinces

[1]Third National Mortality Retrospective Sampling Survey.

accused the upstream provinces of polluting the river water. According to them, the upstream provinces would dump their pollution into the river in order to evacuate it to other provinces. Although China meets all the conditions leading to the emergence of such a problem, to our knowledge no one has studied whether there were transboundary pollution problems in this country. Indeed, most of the studies focus on the United States (Helland and Whitford, 2003; Kahn, 2004; Sigman, 2005; Konisky and Woods, 2009). Thus, this article contributes to the literature by assessing whether there is a problem of transboundary pollution in China.

To do so, we study the location choice of polluting firms between 2002 and 2008 in Hebei, one of the most polluted provinces of the country. Specifically, we test whether polluting firms are more likely to set up, *ceteris paribus*, in counties,[2] that share a border with another province than in interior counties. For this purpose, we use the lists of polluting firms published annually by the Ministry of Environmental Protection (MEP) and by the Environmental Protection Bureau (BEP) of Hebei Province. To ensure the robustness of our results, several measures of the variable of interest, *border*, are constructed from GIS data. Estimations of a count-data model (negative binomial model and zero inflated negative binomial model) allow us to conclude that border counties are more attractive destinations for polluting firms than interior counties. Moreover, it appears that this effect has been strengthening over time.

The rest of the article is organized as follows. Following the introduction, in a second part, we briefly present the Chinese environmental policy and its effect on the polluting enterprises. Next, we study why, in the context of a decentralized policy, polluting enterprises would tend to agglomerate near borders. The fourth part is devoted to the presentation of the data and the model. We present the estimation strategy in a fifth part and results in a sixth part. Finally, we conclude and propose some policy recommendations.

2 Environmental policy in China

2.1 Organization and evolution of the environmental policy

The Chinese environmental policy has gradually developed since the late 1970s, when the first environmental protection laws were adopted. At the beginning, the power of the regional environmental agencies was extremely limited. During the 1980s, the environmental policy followed

[2] The "county" corresponds to the third level of administrative division in China. There are three types of administrative divisions at the county level in China: prefectural city districts (*shìxiaqu*), cities at county level (*xiànjishì*) and the counties (*xiàn*).

the general movement of decentralization occurring in the country ; since then, environmental protection has depended largely on local governments. Nevertheless, it is only after 1990 that environmental protection really becomes a political objective (Sinkule and Ortolano, 1995). The 2000s mark a new step in China's environmental policy, which has clearly been tightened over the past decade. For example, from 2002 to 2008, the total amount of collected pollution levies rose from 6.74 to 18.52 billion yuan (Ministry of Environmental Protection, 2003 ; 2009[3]).

As we said, since the 1980s, the Chinese environmental policy is decentralized. It is thereby managed at the national level by the MEP and at the regional and local level (provinces, prefectures and counties) by the EPBs. The central government (MEP) establishes environmental standards, is responsible for coordinating regional interests and conflicts, and evaluates regional environmental performances. However, environmental policy is implemented by the regions (EPBs). They monitor the emissions of polluters and impose penalties if the standards are not met (OECD, 2006). Decentralization of environmental policy in a country as heterogeneous as China offers undeniable advantages. Indeed, a decentralized environmental policy allows for greater flexibility as well as better information and adaptation to the local context. In other words, a decentralized policy is more efficient than a centralized one that would apply uniform rules across the country. However, Chinese local governors are evaluated more on their economic performance than on their environmental performance (Li and Zhou, 2005). Therefore, environmental protection has often been sacrificed to economic performance.

Facing this critical situation, we have recently observed a certain recentralization of environmental policy in China. In 2008, the State Environmental Protection Agency (SEPA) became the Ministry of Environmental Protection in order to give more power to central government in terms of environmental protection. Moreover, between 2006 and 2008, six major supervision centers [4] were created. Each of these major centers is responsible for several provinces and supervises if they respect the environmental standards established by the central government. The centers are also in charge of coordinating interprovincial conflicts. These new centers constitute in fact a new intermediate level between central government and provinces, created in order to limit the negative effects of decentralization. However, until now, the power of these

[3] State of the environment in China.

[4] Five centers were created in 2006 : the center of South China (Hubei, Hunan, Guangdong, Guangxi and Hainan), of the Southwest (Chongqing, Sichuan, Guizhou, Yunnan and Tibet), of the Notheast (Liaoning Jilin and Heilongjiang), of the Northwest (Shaanxi, Ningxia, Gansu, Qinghai and Xinjiang) and of the East (Shanghai, Jiangsu, Zhejiang, Fujian, Anhui, Jiangxi and Shandong). In 2008 the center of North China was created (Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia and Henan).

new centers has been very limited and the environmental policy is still largely implemented by Chinese provinces.

2.2 Impact of the environmental regulation on polluting firms in China

Several studies estimate that China's environmental policy has an important effect on the polluting firms, forcing them to reduce their emissions and to spend money on the treatment of waste. For example, environmental controls significantly affect the environmental performance of polluting industries in the city of Zhenjiang (Dasgupta *et al.*, 2001). Other studies estimate that the levies have a very significant impact on the level of industrial pollution (Wang and Wheeler, 2003) and on the waste water treatment expenditure (Wang, 2002). Moreover, although state-owned enterprises have more bargaining power with local authorities in terms of charges to pay (Wang *et al.*, 2003), environmental policy also has a significant impact on them (Wang and Wheeler, 2005).

3 Literature review: why would polluting firms be more likely to set up near borders?

Like all plants, those which pollute do not choose their location randomly: they decide to set up in a particular region to maximize their profit[5]. Thus, firms are generally attracted to regions with good market opportunities, where labor is cheap and skilled. Polluting firms could also take into account the severity of environmental regulation. In the Chinese context, where environmental policy is decentralized and where it significantly affects the firms, it is expected that within a province, border counties constitute particularly attractive destinations for polluting firms. Two elements could lead polluting firms to set up near borders: "pollution havens" and "free-riding" effects.

3.1 Differences in interprovincial regulation: *"pollution havens hypothesis"*

In China, the implementation of environmental policy greatly varies from one province to another (Wang and Wheeler, 2005). These regional differences would be at the origin of a "pollution havens" phenomenon[6] within China: polluting firms would be attracted to regions where

[5] Several studies show that in China, the location choice of foreign firms (Wu, 1999; He, 2004) and Chinese firms (Wen, 2004) depends nowadays on "rational economic considerations".

[6] The hypothesis of "pollution havens" is generally considered at the international level. According to this hypothesis, in a world of free trade, the South, whose environmental regulations are less stringent, has a compar-

environmental regulations are less strict (He, 2006; Dean Lovely and Wang, 2009). In this context, it could be very profitable for a firm to locate just at the border between to provinces. Indeed, at the borders, there are discontinuities in environmental regulations (Kahn, 2004). By crossing an administrative boundary, we can move suddenly from strict environmental regulation to a less restrictive one. Crossing a border can therefore be a way to escape from severe environmental regulations while continuing to benefit from the market access of the neighboring province with stricter environmental regulation. Kahn (2004) shows that in the United States, in low environmental regulation states, dirty industries set up more in counties that border high regulation states than in interior counties.

However, this border effect could be limited in China because of a significant inter-provincial protectionism (Poncet, 2003, 2005). Indeed, it would not be profitable to locate in a border county if crossing the border lead to a loss of market access. If the costs associated with the loss of market exceed the benefits generated by lower environmental standards, polluting firms will not settle more in border counties than in those further from the border.

3.2 Differences in intra-provincial regulations: *"free-riding hypothesis"*

When environmental policy is decentralized, regional regulators may implement the policy less tightly in border counties than in interior ones. In other words, free-riding may emerge at the boundaries between different regions of the country[7]. For example, in the U.S., within states, the enforcement of environmental law would be stricter in interior counties than in the counties that border another state (Helland and Whitford, 2003). Several elements could encourage regulators to do this. First of all, at borders, a region's expenditure on pollution control does not solely benefit that region; it is also of benefit to the neighboring regions (Sigman, 2002). Since regions have limited financial resources, they prefer spending them where they can enjoy the full benefit of the spending, that is to say, in interior counties[8]. Secondly, at borders, a part of the pollution of a firm falls on the neighboring region. However, it is usually the region where the company is located which enjoys the overall economic benefits due to the presence of the firm (jobs and taxes) (Helland and Whitford, 2003). Thus, in border counties, the population benefits from the overall positive economic impacts related to the presence of the firm and

ative advantage in producing "dirty" goods. This can lead polluting industries to migrate from the North to the South.

[7] A region refers to a U.S. state or a Chinese province.

[8] It is a positive externality problem: in border counties, expenditure on environmental protection will be below the optimal level.

only suffers from a part of the pollution generated[9]. On the contrary, in interior counties, the population would benefit from job opportunities but would also have to bear all the pollution generated. So, we expect social discontent, related to the set up of a polluting firm, to be higher in interior counties than in border ones. As a result, a regulator concerned with its political support[10] and promotion[11] will be more likely to oppose the arrival of a polluting firm and to apply more stringent environmental regulations in interior counties. Gray and Shadbegian (2004) find that the U.S. plants whose pollution falls partly on the population of neighboring states tend to pollute more. However, according to the authors, this effect would be reduced if the neighboring state had strict environmental standards as it would exert pressure on the polluting state to reduce its level of pollution. Thus, contrary to the "pollution-havens" effect which assumes that polluting firms are concentrated in counties bordering provinces with stricter regulations, according to the "free-riding" effect, polluting firms would agglomerate in counties that border provinces with low regulations.

4 Data and model specification

4.1 Hebei province

In this paper, we study the location decision of polluting firms created between 2002 and 2008 in Hebei Province for several reasons. First of all, Hebei has been industrialized for a long time, which makes it one of the most polluted provinces in the country. For example, according to the list published in 2010 by the Chinese government, which identifies the most polluting firms in China, Hebei is one of the provinces which has the largest number of polluting firms. Among the 9833 top polluters in China, 744 are located in Hebei. Hebei has therefore the highest number of polluting firms just after Jiangsu (838) and Shandong (774). Moreover, Hebei shares

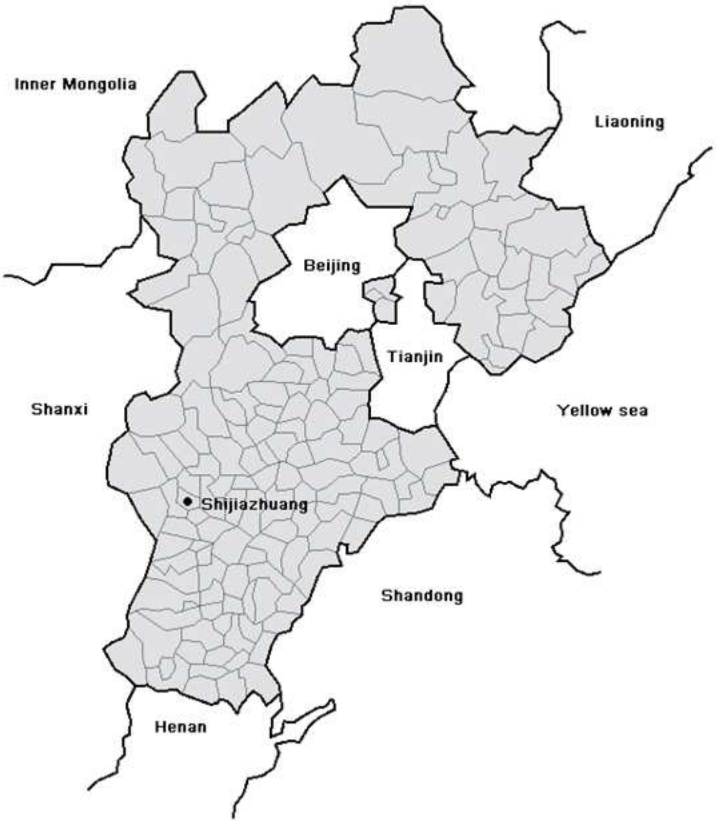
[9]It is a negative externality problem: pollution discharged in border counties will be greater than the optimal level.

[10]The Chinese authorities have to address a large and growing number of citizen complaints about pollution. There were already 138495 letters of complaint in 1993 (Dasgupta and Wheeler, 1997). In 2002, 428626 letters were sent to the authorities (State Environmental Protection Agency, 2003). In some extreme cases, local officials lost their posts because of public pressure after environmental crises. For example, in 2009, numerous local officials were dismissed because of pollution accidents in Hunan, Shaanxi and Inner Mongolia which caused a great amount of intense public pressure.

[11]The political promotion system in China has evoluted over time. While promotion used to be based only on economic performances, from 2005 experiments have been realized in some provinces in which promotion depends both on economic and environmental performances.

borders with seven other provinces including the cities of Beijing and Tianjin. It is interesting to note that Hebei is surrounded by provinces both richer and poorer than itself and therefore, probably more or less stringent than itself in terms of environmental protection[12]. In addition, Hebei has already been involved in several conflicts of transboundary pollution. For example, in August 2010, residents of the city of Chifeng (Inner Mongolia), located downstream from the *Xiluga* river, registered a complaint against a firm in the upstream county of Weichang (Hebei). According to the inhabitants, the firm dumped untreated sewage into the river which would have resulted in the death of their livestock[13]. Finally, Hebei is one of the few provinces whose data are available to carry out our study.

Figure 1: Hebei province: county level divisions



[12]Remind : according to "pollution-havens" effect, polluting firms are concentrated in counties bordering provinces with stricter regulations. On the contrary, according to the "free-riding" effect, polluting firms are more likely to agglomerate in counties bordering provinces with low regulations.

[13]China Environment News, 23 novembre 2010.

4.2 Dependent variable and construction of the sample

The dependent variable of our model is the number of polluting firms created by county[14] between 2002 and 2008. We constructed this variable from the lists published by the MEP and the EPB of Hebei. Since 2007, the MEP and EPBs of provinces annually publish a list that identifies the most polluting firms in China[15]. These lists give the name of each firm and the official code of the county in which it is located. However, no information is given regarding the date of creation of the firm or its level of emissions. So, these lists give the necessary information to estimate a model of stock in which the total number of firms in a county is regressed on a set of regional variables (see Ma (2010) for example). However, to study transboundary pollution effects, we need to estimate a flow model in which the number of firms created in a county at the year t is regressed on the characteristics of this county at year t. Indeed, on the one hand, a firm decides to set up in a county at year t according to the characteristics of the county at year t. Therefore, a flow model seems more suitable. On the other hand, some firms in the lists were created between 1953 and 1978. At this period of time, the set up of a firm did not depend on an economic rationale, unlike that of recent firms (Wen, 2004). As a consequence, it would be impossible to explain with the same model both the location of firms created before the 1980s and of recently established firms. Finally and most importantly, there was no environmental policy in China before 1979, when the first environmental protection law was adopted. To test the existence of border effects, we should take a sample of firms which have been recently created and therefore sensitive to environmental regulation. As a result, in order to estimate a model of flow, we have collected the dates of creation of polluting firms on the official website of the Industrial and Commercial Bureau of Hebei province.

Once the creation dates were obtained, we selected firms created after 2002, year from which we have data for the explanatory variables. In addition, the last list of polluting firms was published in 2010 and it lists the most polluting firms in 2008. Thus, our sample covers the 147 districts of Hebei over the period 2002-2008.

[14]The province of Hebei counts 172 county level divisions: 36 districts, 22 county-level cities and 114 counties. As disaggregated data for the districts of the 11 prefectural-level cities is not available, the districts of a same prefectural-level city are aggregated. That's why our sample is constituted of 147 units at the county level.

[15]The lists identify the most polluting firms at national and provincial level in terms of air, water and sewage pollution. More precisely, the firms identified produce 65% of total industrial emissions of SO2, NOx, COD, NH3-N and heavy metals. As these pollutants can cross the regional borders, this data enables us to test for transboundary pollution. Note that there is a lag of two years between the census of firms in the list and their pollution. Thus, the list of 2007 lists the firms that polluted the most in 2005.

4.3 Variable of interest

Three variables of interest are constructed. Firstly, we follow the literature (Helland and Whitford, 2003; Kahn, 2004; Konisky and Woods, 2009) and construct a dummy variable equal to 1 if the county shares a border with another province, or the sea, and 0 otherwise ($Frontier\_1$). One drawback of this variable is that it does not take into account the length of the common border with another province. Actually, among border counties, some share a very small part of their border with another province while others share more than half of the total length of their border with another province. To take into account the variability among border counties, we create a second variable equal to the length of the common border with another province (or sea) divided by the total length of the border of the county ($Frontier\_2$). The drawback of the first two variables is that they do not take into account the variability between non-border counties. Indeed, among the counties that do not share a border with another province or the sea, some counties are located at the center of the province whereas others are very close to the borders. That is why, we create a third variable equal to the distance between the county seat and the closest border ($Distance$). The two last variables are constructed from with the GIS data of the Chinese Academy of Sciences[16] using ArcGis.

The following tables give some descriptive statistics on the polluting firms in Hebei identified in the lists of the MEP and the BEP. Table 1 gives the stock of firms per 100 km^2 in 2001 and 2008. Non-border counties have a higher number of polluting firms per km^2 than border counties both in 2001 and in 2008. However, the number of polluting firms increased by 40.6% per km^2 in border counties against only 33.64% in non-border counties. Thus, it seems that between 2002 and 2008, polluting firms are more likely to set up in border counties.

Table 1: Stock of firms per km^2 in 2001 and 2008

	N	Firms per 100 km^2 in 2001	Firms per 100 km^2 in 2008	Evolution (%)
All counties	147	0.85	1.15	35.43
Border counties	66	0.54	0.76	40.60
Non-border counties	81	1.10	1.47	33.64

[16]Thematic Database for Human-earth System, Chinese Academy of Sciences.

Table 2: Number of firms created between 2002 and 2008

	N	Total number of firms created	Number of creations per 100 km^2	Number of creations per 100 inhabitants
All counties	147	4.57	0.46	12.68
Border counties	66	7.15	0.51	22.32
non-border counties	81	2.47	0.41	4.82

This result is confirmed by table 2. Between 2002 and 2008, more polluting firms have been created in border counties. This is true no matter whether we take the absolute number of creations, the number of creations per square kilometer or per capita. The differences between data in table 1 (stock of firms) and table 2 (recently created firms) reflect an evolution in the polluting firms choice of location in China. Among the firms identified in the lists, some were created in the 1950s. However, from 1956 to 1978, the location decision did not depend on an economic rational but arose from a strategy aimed at protecting industries from potential destructive military conflicts. Then, from 1965 to 1978, three principles determined the location of industrial firms: "proximity to mountains, dispersion and concealment" (Wen, 2004). Moreover, at that period of time, there was not yet any environmental policy in China. Therefore, it should not be a surprise that the stock of firms is not higher in border counties than in non-border counties. In contrast, newly created polluting firms choose their location according to economic criteria and certainly take into account the degree of environmental policy implementation (see Section 2.2). This would explain why nowadays polluting firms would set up more in border counties than before.

4.4 Other determinants in polluting firms choice of location

As control variables, we introduce the traditional determinants of firms location *i.e.* regional characteristics that may affect firms' profit. First of all, a number of variables affect firms' revenues. On the one hand, firms are attracted to regions with agglomeration economies. By locating where others are located, firms benefit from monetary and non-monetary externalities. Two variables are constructed: agglomeration economies in the industry sector and agglomeration economies in the service sector (the description and definition of variables are given in Appendices 1 and 2). On the other hand, firms are generally attracted to regions that offer significant market opportunities[17]. Generally, firms do not only consider the local market but also the potential external market *i.e.* the markets of neighboring regions (Head and Mayer, 2004). Two variables are used to capture the effects of both local and external markets. The local market is measured by the county's population. The potential (or external) market is a spatial lag variable of the following form:

$$Potential\ market_i = WPOP_i = \sum_j w_{ij} \cdot POP_j$$

with i the county (county, districts or city at county level), j the city (prefectural-level or county-level city). Then, the potential market is measured as the population of neighboring cities weighted by the inverse of the distance between each county i and each city j. Furthermore, as regions whose population is well educated are likely to attract firms, we introduce a variable measuring the average level of education of the population. In China, regions benefiting from special economic zone (SEZ) status also attract significantly more firms (Wu, 1999; Cheng and Stough, 2006). As a result, we introduce a dummy variable equal to the number of SEZs present in the county. Finally, we control by GDP per capita and its squared to check the existence of a Kuznets relationship[18].

On the other side, firms are attracted by regions where production factors are cheap. The more expensive the labor and the land are, the less probability for a region to receive new firms. Thus, we introduce the real wage rate in industry and the population density as a proxy for land price. Note that polluting firms would also prefer areas with low population density where their pollution reaches less people and so, leads to less social discontent. Finally, we introduce dummies to reflect the nature of the administrative unit (county, district of prefectoral cities or city at county level). All of this data comes from Hebei Statistical Yearbooks (2003-2009).

5 Estimation strategy

The dependent variable of the model is the number of firms created in each county between 2002 and 2008. The special nature of the dependent variable (non-negative integers with high frequency of zeros) lead us to estimate a count-data model. The standard model in this case is

[17]Note that the importance of market access depends on production. If a firm produces easily transportable goods or goods for export, market access should not significantly affect its decision.

[18]GDP per capita also controls in part for the level of development of the county and thus, for some variables for which we do not have any information (for example, infrastructures).

the Poisson regression model. This model estimates how much a 1% change in an explanatory variable x_i affects the probability that a firm sets up in the territory i. The probability $Prob(y_i)$ of a territory i to receive y_i firms is based on a set of characteristics x_i of this territory.

$$Prob(y_i) = f(x_i)$$

The most common way to model this probability function is to assume that the variable y_i follows a Poisson distribution. In this case, the probability for a region i to receive y_i firms is given by:

$$Prob(Y = y_i|x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, y = 0, 1, 2, ..., n \text{ et } \ln(\lambda_i) = \beta' X_i$$

The vector of coefficients β is estimated by the method of the maximum likelihood. However, the Poisson model is restrictive because it assumes that λ_i is both the conditional mean and variance of y_i (equi-dispersion hypothesis):

$$Var[y_i|x_i] = E[y_i|x_i] = \lambda_i$$

The hypothesis of equi-dispersion is poorly respected with data on the choice of location of firms as the conditional variance is often higher than the conditional mean (overdispersion). Several phenomena can lead to overdispersion: the clustering of firms due to agglomeration effects (contagion phenomena), the presence of unobserved heterogeneity and an "excess of zeros" (see Appendix 4). In the presence of overdispersion, the Poisson model is no more appropriate because standard deviations obtained are biased and therefore, statistical inferences are invalid. If overdispersion comes from clustering of firms or unobserved heterogeneity, the standard solution is to assume that the variable y_i follows a negative binomial distribution which is more flexible than the Poisson one[19]. The negative binomial model is obtained introducing heterogeneity in the Poisson parameter:

$$\ln(\lambda_i) = \beta' X_i + \varepsilon_i$$

with ε_i following a gamma distribution with mean 1 and variance α. In this model, which is a mixture of Poisson and Gamma distributions, the probability $Prob(y_i)$ of a territory i to receive a number y_i of firms is given by:

[19] If overdispersion comes from an excess of zeros, the solution consists in estimating a zero inflated Poisson model. If overdispersion comes from both clustering of firms and excess of zeros, the appropriate model will be the zero inflated negative binomial model. See Appendix 4 for more details.

$$Prob(Y = y_i) = \frac{\theta^\theta \lambda_i^{y_i}}{\Gamma(y_i+1)\Gamma(\theta)} \frac{\Gamma(y_i+\theta)}{(\lambda_i+\theta)^{y_i+\theta}}$$

where Γ is the gamma function and $\theta = 1/\alpha$, with α the over-dispersion parameter. when $\alpha = 0$, there is equi-dispersion and the negative binomial model is equivalent to the Poisson model (these two models are nested). The vector of the coefficients β is estimated by the method of the maximum likelihood. Note that in the negative binomial model, the variance can be different from the mean as it is linked to the mean by the over-dispersion parameter α which has to be estimated:

$$Var[y_i|x_i] = E[y_i|x_i] \cdot \{1 + \alpha E[y_i|x_i]\} > E[y_i|x_i]$$

Poisson and negative binomial models can be estimated with panel data. Hausman *et al.*, (1984) propose estimators similar to those of the fixed effects and random effects for Poisson and negative binomial models. The Poisson (or negative binomial) model with random effects is asymptotically more efficient than the fixed effects model. However, in the case of correlation between explanatory variables and fixed effects, the random effects model is not consistent. As for linear panel, the lack of correlation between explanatory variables and specific effects can be tested with the Hausman test.

We have annual data over the period 2002-2008. However, we encounter one difficulty with the panel estimation. Indeed, our variable of interest is a geographical variable, by definition, constant over time. To our knowledge, it is impossible to estimate the coefficients associated to non-varying variables when estimating a count data model with fixed-effects (methods of Hausman-Taylor or Plumper and Troeger in the case of linear panels). Moreover, due to the presence of the border variable, it is not accurate to interpret the Hausman test to check whether the random effects model is appropriate. The estimation of the random effects model may therefore produce biased estimates. If the panel estimation meets some difficulties, the cross-sectional one seems appropriate. Indeed, as our variable of interest is a geographical one, it should be exogenous. Then, the cross-sectional estimates seem to be an appropriate alternative to the panel estimation. As a result, to ensure robustness, we run two types of estimations. On the one hand, we run cross-sectional estimations. In this case, the total number of polluting firms created in each county between 2002 and 2008 is regressed on the average value of the explanatory variables of this period. On the other hand, we run random effects estimations. The advantage of the panel estimation is that we can test if border counties have been increasingly attractive for polluting firms over time. Indeed, the tightening of the environmental policy, the increasing number of complaints by the population and the change in the political promotion system could have lead polluting firms to set up more near borders over the last decade.

6 Estimation results and discussion

6.1 Poisson model or negative binomial model ?

The first step consists in checking if the Poisson model fits our data. To this end, we test in several ways whether the hypothesis of equi-dispersion is respected. First of all, the standard deviation of the dependent variable is more than two times its mean, which indicates that there is probably overdispersion (see Appendix 1). Secondly, we estimate the model assuming that the number of created entreprises follows a Poisson distribution[20]. The chi-square value obtained is very high (754.18) indicating that the Poisson distribution is not suitable. Thus, we estimate the model assuming that the dependent variable follows a negative binomial distribution.

6.2 Existence of transboundary pollution problems ?

Tables 3 and 4 present the results. Table 3 presents cross-sectional estimations and table 4 panel estimations. Among the panel estimates, in the first place, we regress the number of firms created at year t on the values of the explanatory variables in t: estimations (4), (5) and (6). Secondly, we regress the number of firms created in t on the explanatory variables in $t-1$ in order to rule out endogeneity: estimations (7), (8) and (9). Furthermore, in each case (cross-section, panel with contemporary variables, panel with lagged variables), we use three different measures of the variable of interest. On the one hand, we use a dummy variable equal to 1 if the county shares a border with another province or with the sea and 0 otherwise (border_1): estimations (1), (4) and (7). On the other hand, we use the ratio of the length of common border over the total length of the county's border (border_2): estimations (2), (5) and (8). The third variable measures the distance between the county seat and the closest border (distance): estimations (3), (6) and (9).

[20]Results available on request.

Table 3: Cross-sectional estimates of the negative binomial model

	(1)	(2)	(3)
Border_1	0.570*		
	(0.059)		
Border_2		2.519***	
		(0.003)	
Distance			-0.331**
			(0.033)
Agglo. industry	-0.322	-0.341	-0.367
	(0.262)	(0.223)	(0.198)
Agglo. service	2.271**	1.986**	2.273***
	(0.011)	(0.015)	(0.008)
Population density	-2.155***	-1.931**	-2.314***
	(0.005)	(0.010)	(0.002)
Wage	-3.210***	-3.241***	-3.406***
	(0.009)	(0.007)	(0.007)
Education	-0.659	-0.431	-0.471
	(0.395)	(0.551)	(0.533)
Local market	1.502***	1.533***	1.599***
	(0.004)	(0.003)	(0.002)
Potential market	-0.556***	-0.507***	-0.529***
	(0.000)	(0.000)	(0.000)
SEZ	0.128	0.0545	0.113
	(0.564)	(0.800)	(0.622)
County	1.977	1.346	1.637
	(0.141)	(0.270)	(0.187)
County level city	1.966	1.331	1.656
	(0.134)	(0.256)	(0.173)
GDP per cap.	0.938**	1.007**	1.025**
	(0.028)	(0.019)	(0.013)
sq. GDP per cap.	-0.0917	-0.107	-0.117
	(0.811)	(0.771)	(0.760)
Constante	31.25***	31.80***	38.15***
	(0.007)	(0.004)	(0.002)
Alpha	0.194	0.119	0.179
	(0.260)	(0.495)	(0.299)
N	147	147	147

P-values in parentheses : * p<0.1 ** p<0.05 *** p<0.01.

Table 4: Random-effects estimates of the negative binomial model

	(4)	(5)	(6)	(7)	(8)	(9)
	Contemporaneous variables			Lagged explanatory variables		
Border_1	0.550**			0.689**		
	(0.032)			(0.017)		
Border_2		1.954***			2.643***	
		(0.003)			(0.000)	
Distance			-0.205*			-0.233*
			(0.096)			(0.083)
Agglo. industry	0.103	0.0999	0.0937	0.140	0.130	0.132
	(0.417)	(0.427)	(0.463)	(0.318)	(0.345)	(0.348)
Agglo. service	1.078**	1.013**	1.056**	0.275	0.230	0.287
	(0.019)	(0.026)	(0.021)	(0.500)	(0.559)	(0.470)
Pop. density	-1.138**	-1.108**	-1.279***	-0.374	-0.365	-0.611
	(0.013)	(0.014)	(0.004)	(0.386)	(0.373)	(0.134)
Wage	-0.834	-0.887*	-0.874*	-1.088*	-1.149**	-1.144*
	(0.108)	(0.085)	(0.093)	(0.061)	(0.046)	(0.050)
Education	0.0126	0.0228	0.0269	-0.239	-0.199	-0.208
	(0.970)	(0.946)	(0.937)	(0.523)	(0.587)	(0.578)
Local market	0.948***	0.967***	1.030***	0.653**	0.701**	0.770**
	(0.002)	(0.001)	(0.001)	(0.040)	(0.023)	(0.015)
Potential market	-0.618***	-0.581***	-0.594***	-0.610***	-0.561***	-0.579***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SEZ	-0.0813	-0.0524	-0.0571	0.145	0.150	0.181
	(0.620)	(0.741)	(0.726)	(0.459)	(0.416)	(0.352)
County	1.529*	1.425	1.408	1.030	0.856	0.907
	(0.094)	(0.114)	(0.124)	(0.286)	(0.365)	(0.345)
County level city	1.308	1.202	1.161	0.979	0.821	0.808
	(0.129)	(0.157)	(0.175)	(0.280)	(0.352)	(0.367)
GDP per cap.	0.419	0.438	0.480*	0.470	0.467	0.524*
	(0.135)	(0.116)	(0.087)	(0.140)	(0.138)	(0.100)
sq. GDP per cap.	-0.499***	-0.504***	-0.500***	-0.405**	-0.430***	-0.417***
	(0.000)	(0.000)	(0.000)	(0.010)	(0.007)	(0.009)
Constante	9.227*	9.535*	12.66**	9.972*	10.40*	14.28**
	(0.067)	(0.054)	(0.013)	(0.077)	(0.060)	(0.013)
a	1.310***	1.344***	1.291***	1.352***	1.422***	1.331***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
b	0.271	0.354*	0.263	0.125	0.234	0.104
	(0.181)	(0.092)	(0.191)	(0.560)	(0.295)	(0.622)
N	1029	1029	1029	882	882	882

P-values in parentheses : * p<0.1 ** p<0.05 *** p<0.01.

In each regression, year dummies are introduced.

In all nine cases, the variable of interest has the expected sign and is statistically significant. Counties that share a (larger part of their) border with another province or with the sea have a higher probability of polluting firms setting up there. In the same way, the further the county seat is from the boundary, the lower the probability of polluting firms settling there. Thus, it seems that in China, within Hebei province, firms are more likely to set up near borders. So, it seems that transboundary pollution problems exist in China. If this result has already been demonstrated for the U.S. case (Helland and Whitford, 2003; Kahn, 2004; Konisky and Woods, 2009), to our knowledge we are the first to demonstrate this phenomenon in China.

Regarding the control variables, their sign and significance are coherent and relatively robust. Counties with a more developed tertiary sector, a lower population density, lower labor costs and a larger local market attract more polluting firms. This confirms that the location choice of Chinese firms nowadays depends on economic factors.

6.3 Has transboundary pollution increased over time?

In the second part of this study, we relax the assumption that being close to the boundary has a constant effect on the location choice of firms over time. To test whether polluting firms increasingly set up in border counties over time, an interactive variable *variable of intesrest*year* is introduced into the model. Table 5 gives the estimation results. We present here only the coefficients associated with the interactive variables *variable of intesrest*year*. The complete table is available in Appendix 3.

As in section 6.2, we perform two types of panel estimations: in estimations (1), (2) and (3), the dependent and the explanatory variables are observed at year t. In estimations (4), (5) and (6), we regress the dependent variable at t on the explanatory variables in $t-1$. According to these estimations, border counties have become increasingly attractive destinations over the period 2002-2008. In the same way, counties located far from borders have become less attractive destinations. Indeed, while during the first years of the sample the variable of interest is not significant or not robust, it becomes significant and its coefficient increases each year. Several elements can explain the increasing attractiveness of border counties. Firstly, as discussed in Section 2.2, environmental policy in China has been tightened since 2000, which could have led to a perverse effect. With the tightening of environmental regulations, it is expected that polluting firms are increasingly sensitive to environmental regulations. Thus, it is likely that firms care more about differences in the application of environmental standards and are more attracted to the border counties over time. Even if the differences in the application of environmental standards between interior and border counties remain the same over time, they could play a more important role today than before. Secondly, the Beijing Olympics in 2008 led to the relocation of a large number of polluting firms. In fact, some polluting firms located in Beijing were closed and re-opened in the neighboring province - Hebei. The Olympic Games have also made the creation of polluting firms in Beijing more difficult. It is possible that firms wishing to set up in Beijing moved to Hebei as no better option was available. As these firms come to Hebei because they cannot set up in Beijing, these firms are expected to set up as close as possible to Beijing *i.e.*,in the counties sharing a border with Beijing. Finally, the increasing number of citizen complaints regarding pollution and the change in the political promotion system could have lead local regulators to harden the implementation of the environmental policy particularly in interior counties.

Table 5: Has transboundary pollution been increasing over time?

	(1)	(2)	(3)	(4)	(5)	(6)
	Contemporaneous variables			Lagged explanatory variables		
	Border_1	Border_2	Distance	Border_1	Border_2	Distance
Var_interest*2002	0.0473	0.262	0.0196	0.160	0.144	0.332*
	(0.890)	(0.773)	(0.916)	(0.649)	(0.880)	(0.092)
Var_interest*2003	0.168	0.201	0.307	0.852**	2.594***	-0.260
	(0.614)	(0.826)	(0.111)	(0.033)	(0.008)	(0.184)
Var_interest*2004	0.838**	2.648***	-0.301	0.294	2.068**	-0.204
	(0.027)	(0.005)	(0.116)	(0.441)	(0.039)	(0.313)
Var_interest*2005	0.324	2.040**	-0.214	1.151***	4.589***	-0.498**
	(0.377)	(0.037)	(0.276)	(0.007)	(0.000)	(0.014)
Var_interest*2006	1.159***	4.490***	-0.501***	1.515***	5.803***	-0.663***
	(0.005)	(0.000)	(0.010)	(0.004)	(0.000)	(0.002)
Var_interest*2007	1.491***	5.797***	-0.634***	1.364***	5.309***	-0.550***
	(0.004)	(0.000)	(0.002)	(0.005)	(0.000)	(0.009)
Var_interest*2008	1.277***	5.144***	-0.524**			
	(0.008)	(0.000)	(0.011)			
N	1029	1029	1029	882	882	882

P-values in parentheses : * p<0.1, ** p<0.05, *** p<0.01.

6.4 Robustness checks

6.4.1 Transboundary pollution or access to international markets?

Following the literature, we considered as border county every county that shares a common border with another province or with the sea. Our results could be biased, however, if polluting firms set up more in coastal counties in order to benefit from a good access to international markets. To check if the results we found are due to transboundary pollution or to international market access, we disaggregate the indicators of interest ($border\_1$ and $border\_2$) into two variables. First, we define as "terrestrial border county" a county that borders another province. Second, we define as "coastal county" a county that borders the Yellow Sea. If polluting firms set up in coastal counties to benefit from a good access to international markets, the coefficient associated to the variable *terrestrial border* should not be significant. The results of the estimations, given in Table 6, validate the transboundary pollution hypothesis. Indeed, in five of the six estimations, the coefficient associated to the variable *terrestrial border* is positive and statistically significant. Counties that share a (larger part of their) border with another province have a higher probability of polluting firms setting up there. This confirms that transboundary pollution does exist in China.

Table 6: Transboundary pollution or access to international markets?

	Cross-section		**Panel**			
			Contemporaneous variables		Lagged explanatory variables	
	Border_1	Border_2	Border_1	Border_2	Border_1	Border_2
Terrestrial border	0.565*	2.628***	0.544	1.667**	0.877**	2.396***
	(0.077)	(0.006)	(0.144)	(0.017)	(0.034)	(0.002)
Coastal county	0.665	1.729	1.589**	4.069**	2.223***	4.786**
	(0.157)	(0.312)	(0.020)	(0.020)	(0.004)	(0.013)
Agglo. industry	-0.303	-0.353	0.105	0.107	0.150	0.139
	(0.278)	(0.205)	(0.404)	(0.395)	(0.278)	(0.310)
Agglo. service	2.133**	2.014**	1.071**	0.992**	0.220	0.203
	(0.017)	(0.015)	(0.021)	(0.030)	(0.582)	(0.605)
Pop. density	-1.984**	-1.952**	-1.100**	-1.058**	-0.228	-0.305
	(0.010)	(0.011)	(0.017)	(0.019)	(0.596)	(0.458)
Wage	-3.056**	-3.400***	-0.650	-0.683	-0.832	-0.959
	(0.017)	(0.009)	(0.213)	(0.194)	(0.156)	(0.105)
Education	-0.720	-0.501	0.0570	0.0744	-0.185	-0.150
	(0.360)	(0.504)	(0.867)	(0.826)	(0.623)	(0.687)
Local Market	1.496***	1.538***	0.999***	1.001***	0.707**	0.731**
	(0.003)	(0.003)	(0.001)	(0.001)	(0.025)	(0.018)
Potential Market	-0.571***	-0.496***	-0.654***	-0.615***	-0.661***	-0.594***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SEZ	0.0861	0.0352	-0.0614	-0.0359	0.128	0.164
	(0.705)	(0.872)	(0.712)	(0.822)	(0.516)	(0.376)
County	1.954	1.301	1.717*	1.570*	1.252	1.002
	(0.139)	(0.278)	(0.061)	(0.085)	(0.195)	(0.293)
County level city	1.814	1.281	1.376	1.327	1.033	0.943
	(0.158)	(0.266)	(0.110)	(0.123)	(0.248)	(0.288)
GDP per cap.	0.929**	1.069**	0.297	0.312	0.306	0.340
	(0.033)	(0.018)	(0.312)	(0.290)	(0.353)	(0.306)
sq. GDP per cap.	-0.0726	-0.0805	-0.488***	-0.514***	-0.392**	-0.439***
	(0.855)	(0.826)	(0.000)	(0.000)	(0.012)	(0.006)
Constante	29.07**	33.12***	7.243	7.437	6.787	8.372
	(0.011)	(0.005)	(0.157)	(0.147)	(0.240)	(0.146)
Alpha	0.176	0.115				
	(0.324)	(0.510)				
a			1.332***	1.348***	1.386***	1.432***
			(0.000)	(0.000)	(0.000)	(0.000)
b			0.280	0.334	0.128	0.219
			(0.171)	(0.112)	(0.553)	(0.329)
N	147	147	1029	1029	882	882

P-values in parentheses : * p<0.1 ** p<0.05 *** p<0.01.

In each panel regression, year dummies are introduced.

6.4.2 "Zero inflation problem"

Count-data models allow for estimating models where the dependent variable takes the value zero many times. However, sometimes the dependent variable y_i takes the value zero more times that assumed by the Poisson or the negative binomial distribution. This is known as a "zero inflation problem". This problem arises when two different phenomena lead the dependent variable to take the value zero (see Appendix 4). An excess of zeros may lead to biased results. In this case, the appropriate model to be estimated is a zero-inflated model.

In Hebei province from 2002 to 2008, no polluting firms settled in 36 counties, which represents nearly 25% of the sample. The frequency of zeros in our sample is relatively low compared with studies estimating a zero-inflated model (List, 2001; Roberto, 2004).

Table 7: Distribution of firms creations

Number of creations	0	1	2	3	4	5	> 5
Frequency	36	43	24	12	8	3	21
Percentage	24.49	29.25	16.33	8.16	5.44	2.04	14.29

N.B.: the table gives the frequency and the percentage of counties with 0, 1, 2, ... , creations of firms from 2002 to 2008.

We test the presence of zero inflation with the Vuong test (1989). This test enables us to check if the zero-inflated negative binomial model (ZINBM) is more approriate than the negative binomial model[21] (NBM). In other words, we test whether the previously observed over-dispersion comes only from agglomeration phenomena and/or unobserved heterogeneity (NBM) or if it also arises from zero inflation (ZINBM). The following table gives the values obtained from the Vuong test.

Table 8: Vuong test

	Cross-section	**Panel**	
Variable of interest		Contemporary variables	Lagged explanatory variables
Border1	No convergence	-2.46	-2.02
Border2	No convergence	-2.28	-1.47
Distance	No convergence	-2.61	-1.81

The model estimated in the first step is a logit model

[21] As discussed in Section 4.1., the negative binomial model is prefered to the Poisson one as there is over-dispersion in our data due either to the agglomeration phenomena or to unobserved heterogeneity.

The Vuong test does not favor in any case the zero inflated model. Indeed, the value obtained is never higher than 1.96[22]. The test clearly rejects the ZINBM in favor of the NBM when the model is estimated in panel with contemporaneous variables. However, it fails to conclude in favor of one of the two models when the model is estimated in panel with lagged explanatory variables. As a consequence, we present the estimation results of the ZINBM model in panel with lagged variables. According to table 8, taking into account the potential problem of zero-inflation does not change the previous results as the variables of interest are still of the expected sign and significant.

[22] Asymptotically, the Vuong test statistic has a standard normal distribution and hence, the test statistic obtained has to be compared with the critical value of the normal distribution (1.96). A value above 1.96 (below -1.96, respectively) rejects the negative binomial model (zero-inflated negative binomial model) in favor of the zero-inflated negative binomial model (negative binomial model). Note: For panel data, this test exists only in pooling.

Table 9: Estimations of the zero-inflated negative binomial model

	(1)	(2)	(3)
Border_1	0.664***		
	(0.005)		
Border_2		2.896***	
		(0.000)	
Distance			-0.112**
			(0.016)
Agglo. industry	-0.013	-0.034	-0.044
	(0.909)	(0.764)	(0.742)
Agglo. service	1.136***	1.004**	1.271***
	(0.003)	(0.011)	(0.002)
Pop. density	-1.209***	-1.091***	-1.470***
	(0.001)	(0.004)	(0.000)
Wage	-1.195*	-1.158*	-1.153
	(0.077)	(0.076)	(0.123)
Education	-0.197	-0.087	-0.141
	(0.638)	(0.838)	(0.773)
Local market	0.737***	0.801***	0.862***
	(0.010)	(0.005)	(0.005)
Potential market	-0.433***	-0.381***	-0.426***
	(0.000)	(0.000)	(0.000)
SEZ	0.278	0.256	0.349
	(0.173)	(0.167)	(0.131)
County	1.527*	1.126	1.675*
	(0.075)	(0.179)	(0.075)
County level city	1.728**	1.379*	1.949**
	(0.028)	(0.071)	(0.024)
GDP per cap.	0.063	-0.013	0.016
	(0.831)	(0.964)	(0.962)
sq. GDP per cap.	-0.367	-0.413*	-0.423
	(0.114)	(0.090)	(0.103)
Constante	12.102*	11.594*	13.252*
	(0.057)	(0.062)	(0.061)
Alpha	1.083***	0.843***	1.005***
	(0.000)	(0.000)	(0.000)
Tau	-0.459**	-0.399**	-0.278
	(0.028)	(0.045)	(0.106)
N	882	882	882

P-values in parentheses : * p<0.1 ** p<0.05 *** p<0.01. Year dummies introduced in each regression.

6.5 Discussion of the results

According to our estimations, transboundary pollution problems exist in China. Indeed, in Hebei Province, counties that share a border with another province or the sea have a higher probability, *ceteris paribus*, of attracting polluting firms. Inversely, counties far from borders have a lower probability of seeing polluting firms set up in their territory. Moreover, it appears that this effect has increased over time. However, this last result should be interpreted with caution. To check if border counties become increasingly attractive over time, we run random effects estimations without being able to check the lack of correlation between the fixed effects and the explanatory variables with the Hausman test. More generally, we should be cautious in interpreting our results because the attractiveness of border counties could come from other elements than (intra-and inter-provincial) environmental regulation differences. Indeed, on the one hand, at borders, there are not only environmental discontinuities but also many other types of regulation discontinuities. For example, Holmes (1998) distinguishes in the U.S. ”probusiness” states, which have a right-to-work law (no union shop), from ”anti-business” states. In this case, crossing a border and setting up in a probusiness state, enables a manufacturer to benefit from a more favorable state policy as it weakens unions. The author estimates that in the US, manufacturing activity significantly increases when one crosses the border from an anti-business state into a probusiness state. On the other hand, as suggested by Holmes (1998), sometimes borders are also made by geographical discontinuities. For example, it is not unusual that a mountain range or a river mark the border between two regions or countries. In our study, the only ”natural” border we find is that between Hebei and Shandong: the Canal of Wei (*Weiyunhe*) and the New River of Zhangwei (*Zhangweixinhe*) mark the border between these two provinces. In this case, polluting firms could set up in border counties along the river to benefit from the access to cheap water. As Hebei province shares border with seven provinces, we expect that this should not biase our results.

7 Conclusion

In this paper, we study whether there is a problem of transboundary pollution in China. To do so, we estimate whether polluting firms set up more in counties close to the regional border. Indeed, in the Chinese context where environmental policy is decentralized and where it significantly affects polluting industries, it is likely that a transboundary pollution problem exists. To study the location decisions of polluting firms in Hebei province, we estimate a negative binomial model. Our estimation results suggest that the closer a county is to the border, the higher the probability of it attracting polluting firms. In addition, it appears that this effect has strengthened over time. Thus, there is a risk that people in border counties suffer disproportionately from pollution. Kahn (2004) shows that in the United States, cancer rates are substantially higher in border counties than in interior counties because of transboundary pollution problems.

If transboundary pollution problems are often put forward by opponents of decentralization, our results do not suggest that a centralized policy would be optimal. Indeed, a decentralized policy offers compelling advantages in a country as heterogeneous as China. While a centralized policy would consist in applying uniform rules across the country, a decentralized policy allows for adaption to the local conditions and so, is more efficient. It is unclear which of these two types of policy would lead to higher social welfare. Thus, as suggested by Sigman (2005) in the case of the United States, the optimal policy might be to provide targeted solutions to transboundary pollution problems while staying within the framework of a decentralized policy. The recent creation of the six major regional centers could be a way to reduce these perverse effects. For the moment, the creation of these centers is too recent and their power is still too limited to offer measurable effects. It could be interesting to study the location choices of firms in the period to come to test whether the creation of these intermediate poles, between central government and regional governments, can be a solution to the transboundary pollution problem. It could also be interesting to analyse other aspects unsolved in this study. For example, one could test if transboundary pollution problems are due to a free-riding phenomenon and/or to a pollution haven phenomenon. One could also make a more detailed analysis of firms' location choices according to their characteristics, such as type of pollutant discharged or the ownership regime.

References

Cost of pollution in china: Economic estimates of physical damages. Tech. rep., World Bank, (2007).

From poor areas to poor people: China's evolving poverty reduction agenda, an assessment of poverty and inequality in china. Tech. rep., World Bank, (2009).

Cheng, S., and Stough, R. Location decisions of japanese new manufacturing plants in china: a discrete-choice analysis. *The Annals of Regional Science 40*, 2 (2006), 369–387.

Dasgupta, S., and Wheeler, D. Citizen complaints as environmental indicators : evidence from china. Policy Research Working Paper Series 1704, The World Bank, (1997).

Dean, J. M., Lovely, M. E., and Wang, H. Are foreign investors attracted to weak environmental regulations? evaluating the evidence from china. *Journal of Development Economics 90*, 1 (2009), 1–13.

Gabe, T. Local industry agglomeration and new business activity. *Growth and Change 34*, 1 (2003), 17–39.

Gray, W. B., and Shadbegian, R. J. 'optimal' pollution abatement–whose benefits matter, and how much? *Journal of Environmental Economics and Management 47*, 3 (2004), 510–534.

Greene, W. H. Accounting for excess zeros and sample selection in poisson and negative binomial regression models. Working Papers 94-10, New York University, Leonard N. Stern School of Business, Department of Economics, (1994).

Hausman, J., Hall, B. H., and Griliches, Z. Econometric models for count data with an application to the Patents-R&D relationship. *Econometrica 52*, 4 (1984), 909–38.

He, C. Location of foreign manufacturers in china: Agglomeration economies and country of origin effects. *Papers in Regional Science 82*, 3 (2003), 351–1.

He, J. Pollution haven hypothesis and environmental impacts of foreign direct investment: The case of industrial emission of sulfur dioxide (SO2) in chinese provinces. *Ecological Economics 60*, 1 (2006), 228–245.

Head, K., and Mayer, T. Market potential and the location of japanese investment in the european union. *Review of Economics and Statistics 86*, 4 (2004), 959–972.

HELLAND, E., AND WHITFORD, A. B. Pollution incidence and political jurisdiction: evidence from the TRI. *Journal of Environmental Economics and Management 46*, 3 (2003), 403–424.

HOLMES, T. J. The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy 106*, 4 (1998), 667–705.

KAHN, M. E. Domestic pollution havens: evidence from cancer deaths in border counties. *Journal of Urban Economics 56*, 1 (2004), 51–69.

KONISKY, D. M., AND WOODS, N. D. Exporting air pollution? regulatory enforcement and environmental free riding in the united states. *Political Research Quaterly* (2009).

LI, H., AND ZHOU, L. Political turnover and economic performance: the incentive role of personnel control in china. *Journal of Public Economics 89*, 9-10 (2005), 1743–1762.

LIST, J. A. US county-level determinants of inbound FDI: evidence from a two-step modified count data model. *International Journal of Industrial Organization 19*, 6 (2001), 953–973.

MA, C. Who bears the environmental burden in China–An analysis of the distribution of industrial pollution sources? *Ecological Economics 69*, 9 (2010), 1869–1876.

OCDE. Environmental compliance and enforcement in china-an assessment of current practices and ways forward, (2006).

PONCET, S. Measuring chinese domestic and international integration. *China Economic Review 14*, 1 (2003), 1–21.

PONCET, S. A fragmented china: Measure and determinants of chinese domestic market disintegration. *Review of International Economics 13*, 3 (2005), 409–430.

ROBERTO, B. Acquisition versus greenfield investment: the location of foreign manufacturers in italy. *Regional Science and Urban Economics 34*, 1 (2004), 3–25.

SIGMAN, H. International spillovers and water quality in rivers: Do countries free ride? *The American Economic Review 92*, 4 (2002), 1152–1159.

SIGMAN, H. Transboundary spillovers and decentralization of environmental policies. *Journal of Environmental Economics and Management 50*, 1 (2005), 82–101.

SINKULE, B. J., AND ORTOLANO, L. *Implementing environmental policy in China.* Greenwood Publishing Group, (1995).

Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 2 (1989), 307–33.

Wang, H. Pollution regulation and abatement efforts: evidence from china. *Ecological Economics 41*, 1 (2002), 85–94.

Wang, H., Mamingi, N., Laplante, B., and Dasgupta, S. Incomplete enforcement of pollution regulation: Bargaining power of chinese factories. *Environmental & Resource Economics 24*, 3 (2003), 245–262.

Wang, H., and Wheeler, D. Equilibrium pollution and economic development in china. *Environment and Development Economics 8*, 03 (2003), 451–466.

Wang, H., and Wheeler, D. Financial incentives and endogenous enforcement in china's pollution levy system. *Journal of Environmental Economics and Management 49*, 1 (2005), 174–196.

Wen, M. Relocation and agglomeration of chinese industry. *Journal of Development Economics 73*, 1 (2004), 329–347.

Wu, F. Intrametropolitan FDI firm location in guangzhou, china a poisson and negative binomial analysis. *The Annals of Regional Science 33*, 4 (1999), 535–555.

APPENDIX 1 : DESCRIPTIVE STATISTICS

Variable	Obs	Mean	Standard deviation	Min	Max
Creation of firms	147	4.57	10.08	0	56
Agglomeration in industry	147	29.19	94.58	0.15	573.37
Agglomeration in services	147	38.17	97.94	1.47	643.3
Average salary	147	11634.6	2174.87	6903.57	20822.44
Population density	147	683.23	798.49	43.87	5212.67
Education	147	7.11	1.42	3.38	11.99
Local market	147	467946.4	324134.6	113000	2991172
Potential market	147	63.6	25.22	0	132.51
Special Economic Zones	147	0.3	0.62	0	4.86
Border_1	147	0.45	0.5	0	1
Border_2	147	14.36	20.59	0	74.28
Distance	147	40631.43	25710.47	354.40	110647.60
GDP pc	147	12210.33	7337.96	3775.31	37620.39
County	147	0.78	0.42	0	1
County city	147	0.15	0.36	0	1

APPENDIX 2 : DEFINITION OF VARIABLES

Variable	Definition	Unity
Creation of firms	Number of creations of polluting firms between 2002 and 2008	Création
Agglomeration in industry	Number of employed persons in industry per km^2	Persons per km^2
Agglomeration in services	Number of employed persons in services per km^2	Persons per km^2
Average salary	Average real salary (2002 price)	Yuan
Population density	Population per km^2	Persons per km^2
Education	Part of secondary students in the total population	%
Local market	Total population	Person
Potential market	Population of neighbor cities weighted by distance	
Special Economic Zones	Number of SEZ	SEZ
Border_1	Dummy equal to 1 if the county share a border with the sea or an other province, 0 if not	
Border_2	Length of the common border (with another province or the sea) divided by the total length of the county's boundary	%
Distance	Distance between the county seat and the closest border (with another province or the sea)	meters
GDP pc	Real GDP per capita (2002 price)	Yuan
County	Dummy equal to 1 if county, 0 if not	
County city	Dummy equal to 1 if city at county level, 0 if not	

APPENDIX 3 : EVOLUTION OF BORDER EFFECTS OVER TIME

	(1)	(2)	(3)	(4)	(5)	(6)
	Contemporaneous variables			Lagged explanatory variables		
	Border_1	Border_2	Distance	Border_1	Border_2	Distance
Agglo. industry	0.127	0.107	0.144	0.149	0.118	0.139
	(0.323)	(0.401)	(0.270)	(0.287)	(0.389)	(0.326)
Agglo. service	1.249***	1.091**	1.076**	0.411	0.192	0.213
	(0.008)	(0.017)	(0.021)	(0.356)	(0.631)	(0.653)
Pop. density	-1.296***	-1.174***	-1.351***	-0.484	-0.312	-0.571
	(0.006)	(0.010)	(0.003)	(0.298)	(0.461)	(0.232)
Wage	-0.967*	-1.084**	-0.990*	-1.213**	-1.319**	-1.233**
	(0.067)	(0.034)	(0.058)	(0.038)	(0.021)	(0.035)
Education	0.138	0.384	0.181	-0.140	0.0723	-0.109
	(0.697)	(0.288)	(0.601)	(0.710)	(0.850)	(0.770)
Local market	0.972***	0.968***	1.040***	0.674**	0.653**	0.732**
	(0.002)	(0.002)	(0.001)	(0.038)	(0.039)	(0.026)
Potential market	-0.576***	-0.566***	-0.574***	-0.570***	-0.535***	-0.546***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SEZ	-0.0864	0.00272	-0.0369	0.160	0.240	0.201
	(0.600)	(0.987)	(0.824)	(0.414)	(0.211)	(0.312)
County	1.762*	1.497	1.559*	1.221	0.809	0.802
	(0.059)	(0.108)	(0.096)	(0.223)	(0.405)	(0.437)
County level city	1.743*	1.534*	1.381	1.361	1.003	0.854
	(0.051)	(0.081)	(0.117)	(0.154)	(0.267)	(0.377)
GDP per cap.	0.160	0.141	0.342	0.243	0.252	0.440
	(0.598)	(0.632)	(0.244)	(0.471)	(0.438)	(0.188)
sq. GDP per cap.	-0.374***	-0.241*	-0.368***	-0.267*	-0.126	-0.243
	(0.008)	(0.079)	(0.009)	(0.094)	(0.431)	(0.131)

(End Annexe 3)

	(1)	(2)	(3	(4)	(5)	(6)
Var_interest*2002	0.0473	0.262	0.0196	0.160	0.144	0.332*
	(0.890)	(0.773)	(0.916)	(0.649)	(0.880)	(0.092)
Var_interest*2003	0.168	0.201	0.307	0.852**	2.594***	-0.260
	(0.614)	(0.826)	(0.111)	(0.033)	(0.008)	(0.184)
Var_interest*2004	0.838**	2.648***	-0.301	0.294	2.068**	-0.204
	(0.027)	(0.005)	(0.116)	(0.441)	(0.039)	(0.313)
Var_interest*2005	0.324	2.040**	-0.214	1.151***	4.589***	-0.498**
	(0.377)	(0.037)	(0.276)	(0.007)	(0.000)	(0.014)
Var_interest*2006	1.159***	4.490***	-0.501***	1.515***	5.803***	-0.663***
	(0.005)	(0.000)	(0.010)	(0.004)	(0.000)	(0.002)
Var_interest*2007	1.491***	5.797***	-0.634***	1.364***	5.309***	-0.550***
	(0.004)	(0.000)	(0.002)	(0.005)	(0.000)	(0.009)
Var_interest*2008	1.277***	5.144***	-0.524**			
	(0.008)	(0.000)	(0.011)			
Constante	11.04**	12.80**	11.96**	11.60**	13.25**	9.811*
	(0.033)	(0.011)	(0.023)	(0.043)	(0.018)	(0.090)
a						
	1.405***	1.603***	1.400***	1.476***	1.699***	1.501***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
b						
	0.210	0.216	0.187	0.0708	0.0963	0.0135
	(0.305)	(0.286)	(0.347)	(0.743)	(0.653)	(0.948)
N	1029	1029	1029	882	882	882

P-values in parenthesis : * p<0.1, ** p<0.05, *** p<0.01. In each regression, year dummies are introduced.

APPENDIX 4 : ZERO INFLATED MODEL

An excess of zeros can lead to overdispersion in the data and result in biased estimations (List, 2001). The problem of zero inflation arises when the absence of creation of polluting firms arise through two separate processes (Greene, 1994). In our case, two phenomena could explain why no polluting firms set up in several counties in Hebei during the sample period. First, there is a group of counties in which no polluting firms will never set up, regardless of circumstances, because of their characteristics (underdevelopped local market, unskilled workers...). These counties always attract zero polluting firms; they are ”structurally unattractive”. On the other hand, there are other counties for which the number of new polluting firms might follow a negative binomial regression. However, these counties did not attract any new polluting firms during the sample period. The solution in this case is to estimate a two-stage model. On the one hand, a binary model (logit or probit) is used to distinguish the ”structurally unattractive” territories from other territories. On the other hand, a Poisson (or negative binomial) model is used to explain the number of firms created in the territories, excluding those that are ”structurally unattractive”. This model is called the zero-inflated Poisson model (ZIP) or zero-inflated negative binomial model (ZINBM). The zero-inflated Poisson model is given by :

$$y_i = 0 \text{ with the probability } P_i$$

$$y_i \rightsquigarrow \text{ poisson}(\lambda_i) \text{ with the probability} 1 - P_i$$

Then, we obtain :

$$Prob[Y_i = 0] = P_i + [1 - P_i]R_i(0)$$

$$Prob[Y_i = y_i | Y > 0] = [1 - P_i]R_i(no\ 0)$$

Where P_i represents the state of probability and R_i the Poisson distribution for the variable y_i : $R_i(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$. To sum up, four different count data models exist depending on the existence of over-dispersion and its origin. The following table presents the models to choose according to the situation met.

Table 10: Summary table of the estimators

	No zero inflation	Zero inflation
Neither contagion effect nor unobserved heterogeneity	Poisson model	Zero-inflated Poisson model
Contagion effects and/or unobserved heterogeneity	Negative binomial model	Zero-inflated negative binomial model

To test whether there is a problem of excess zeros, we must test if the ZIP model (ZINBM) is more appropriate than the Poisson (negative binomial) one. Since the Poisson (negative binomial) and the ZIP (ZINBM) models are not nested, we cannot use standard tests. The solution is to use the Vuong (1989) test.